Comparative Chloroplast Genomes of Pinaceae: Insights into the Mechanism of Diversified Genomic Organizations

Chung-Shien Wu[1,†], Ching-Ping Lin[1,†], Chi-Yao Hsu[1], Rui-Jiang Wang[2], and Shu-Miaw Chaw*[,1]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan

[2]South China Botanical Garden, the Chinese Academy of Sciences, Guangzhou, China

†These authors contributed equally to this work

*Corresponding author: E-mail: smchaw@sinica.edu.tw.

Accepted: 6 March 2011

Abstract

Pinaceae, the largest family of conifers, has diversified organizations of chloroplast genomes (cpDNAs) with the two typical inverted repeats (IRs) highly reduced. To unravel the mechanism of this genomic diversification, we examined the cpDNA organizations from 53 species of the ten Pinaceous genera, including those of *Larix decidua* (122,474 bp), *Picea morrisonicola* (124,168 bp), and *Pseudotsuga wilsoniana* (122,513 bp), which were firstly elucidated. The results uncovered four distinct cpDNA forms (A–C and P) that are due to rearrangements of two ~20 and ~21 kb specific fragments. The C form was documented for the first time and the A form might be the most ancestral one. In addition, only the individuals of *Ps. macrocarpa* and *Ps. wilsoniana* were detected to have isomeric cpDNA forms. Three types (types 1–3) of Pinaceae-specific repeats situated nearby the rearranged fragments were found to be syntenic. We hypothesize that type 1 (949 ± 343 bp) and type 3 (608 ± 73 bp) repeats are substrates for homologous recombination (HR), whereas type 2 repeats are likely inactive for HR because of their relatively short sizes (151 ± 30 bp). Conversions among the four distinct forms may be achieved by HR and mediated by type 1 or 3 repeats, thus resulting in increased diversity of cpDNA organizations. We propose that in the Pinaceae cpDNAs, the reduced IRs have lost HR activity, then decreasing the diversity of cpDNA organizations, but the specific repeats that the evolution endowed Pinaceae complement the reduced IRs and increase the diversity of cpDNA organizations.

Key words: Pinaceae, chloroplast genome, isoform, repeat, structural evolution.

Introduction

Chloroplasts are the plant organelles where photosynthesis takes place. Each chloroplast has its own genome (cpDNA) with a typically circular organization (Palmer 1991). CpDNAs of most land plants consist of four parts, including two copies of large inverted repeats (IRs) separated by a large single copy (LSC) and a small single copy (SSC) region. The core of IRs encodes four ribosomal RNAs (16S, 23S, 4.5S, and 5S), which are believed to have preserved the genomic feature of ancestral cyanobacteria (Tomioka and Sugiura 1983; Turmel et al. 1999). Palmer and Thompson (1982) hypothesized that the presence of IRs might have advantages for maintaining conserved gene orders in cpDNAs. As well, active IR-mediated homologous recombination (HR) might consume most of recombinases, thus resulting in insufficient recombinases for recombination at the LSC and SSC regions (Palmer 1991). However, cpDNAs with conserved IRs but a highly rearranged LSC have been found in a number of diverse lineages such as sunflowers (Kim et al. 2005), Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011), jasmines (Lee et al. 2007), *Trachelium* (Haberle et al. 2008), and gnetophytes (McCoy et al. 2008; Wu et al. 2009). Therefore, some factors other than IRs might also influence the structural evolution of cpDNAs.

Pombert et al. (2005, 2006) discovered a positive correlation between the abundance of short repeats (more than 30 bp) and the degree of rearrangements in the cpDNAs of green algae. And in some vascular plants, such as *Oryza* (Shimada and Sugiura 1989), *Pseudotsuga* (Tsai and Strauss 1989), *Abies* (Tsumura et al. 2000), Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011), Fabaceae (Cai et al. 2008), and *Trachelium* (Haberle et al. 2008), short repeats are usually present near the rearranged fragments of restructured cpDNAs. Indeed, cpDNA transgenic experiments have shown that repeats usually more than 200 bp are effective substrates for HR

(see review of Day and Madesis 2007) though small inversions (5–50 bp) from <25 bp repeat-mediated HR have been found in some angiosperm cpDNAs (Kim and Lee 2005). Comparisons of the legume cpDNAs revealed more abundant repeats in the IR-lacking (IRL) than in IR-containing (IRC) clades (Saski et al. 2005; Cai et al. 2008). In contrast, in the cpDNAs of Geraniaceae, short repeats are less abundant in IRL (e.g., *Erodium*) than in IRC (e.g., *Geranium* and *Pelargonium*) clades (Guisinger et al. 2011). Whether a negative/positive association between the presence of IRs and the number of repeats exists in cpDNAs is still unclear.

Pinaceae, the largest family of conifers, comprises approximately 250 species in ten genera—*Abies*, *Cathaya*, *Cedrus*, *Keteleeria*, *Larix*, *Picea*, *Pinus*, *Pseudolarix*, *Pseudotsuga*, and *Tsuga* (see review by Lin et al. 2010). Differing from the cpDNAs of IRL legumes with their complete loss of a copy of IRs, those of Pinaceae have preserved a rather reduced pair of IRs (236–495 bp) containing only the *3′psbA* and *trnI-CAU* genes (Tsudzuki et al. 1992; Lin et al. 2010). Loss of one IR has been considered robust support for the monophyly of conifers (Raubeson and Jansen 1992). Actually, the complete cpDNA of *Cryptomeria japonica*, a non-Pinaceae conifer, possesses a pair of reduced IRs containing only the *trnI-CAU* gene at a different position from those of Pinaceae cpDNAs (Hirao et al. 2008). Whether this difference connotes an independent evolution of IR reduction between Pinaceae and *Cryptomeria* requires close scrutiny and examination with more non-Pinaceae cpDNAs.

The cpDNAs of Pinaceae are characterized by diversified organizations and many repeats (Hipkins et al. 1994). Strauss et al. (1988) speculated that in the Pinaceae cpDNAs, rearrangements might have occurred after IR reduction and be associated with short repeats. A 40- to 50-kb inversion that distinguishes *Pseudotsuga* from *Pinus* was found to be associated with a pair of 482-bp repeats (Tsai and Strauss 1989). Interestingly, Tsumura et al. (2000) documented that in different populations of both *Abies* and *Tsuga,* two isomeric cpDNAs (named types A and B by the authors) are distinguished from each other by a 42-kb inversion polymorphism. These data led us to ask two questions: Are extensive rearrangements common in the ten Pinaceae genera? And is there a mechanism regulating the diversity of cpDNA organizations? Answering these questions requires a broader sampling across all Pinaceae genera. Therefore, we examined cpDNA organizations in 53 species sampled from the ten Pinaceous genera, including three first-elucidated complete cpDNAs that represent the three Pinaceous genera (*Larix*: *L. decidua*, *Picea*: *Pic. morrisonicola*, and *Pseudotsuga*: *Ps. wilsoniana*). The cpDNA-based structural comparisons among representative species from seven different Pinaceae genera are also presented.

Materials and Methods

Sample Collection and DNA Extraction

Two grams of young leaves for DNA extraction were harvested from 2-year-old seedlings of *L. decidua*, *Pic. morrisonicola*, and *Ps. wilsoniana* in the greenhouse of Academia Sinica. Seeds of the 28 Pinaceae species (table 1) were purchased from Sheffield's Seed Co., USA. The seeds were mixed and stratified in moist peat moss at 4 °C for 30 days to overcome seed dormancy before sowing. One-month-old seedlings were harvested for DNA extraction. Total DNAs of all Pinaceae species were extracted by use of a 2 × CTAB protocol (Stewart and Via 1993).

Long-Range PCR Amplification and Sequencing

Specific cpDNA fragments of *L. decidua*, *Pic. morrisonicola*, and *Ps. wilsoniana* were amplified by long-range polymerase chain reaction (PCR) (TaKaRa LA *Taq*; Takara Bio Inc.) with use of previously published primers (Lin et al. 2010). We covered the entire cpDNA with approximate 12 partially overlapped PCR fragments, which were approximately 6–16 kb long. Each fragment was sequenced by combining at least three independent PCR amplicons to reduce any potential PCR artifact. Amplicons were purified, hydrosheared, cloned, sequenced, and assembled following the method of Wu et al. (2007).

Gene Annotation

Protein-coding and ribosomal RNA genes were annotated by use of DOGMA (http://dogma.ccbb.utexas.edu/), and tRNA genes were predicted by tRNAscan (http://lowelab.ucsc.edu/tRNAscan-SE/).

Dot-Plot Analyses

Dot-plot analyses involved use of a Blast program of the "Align two sequences using Blast (bl2seq)" available at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome).

CpDNA-Wide Local Multiple Alignments

Sequences of the seven complete Pinaceae cpDNAs (*Cedrus deodara* [AB480043], *Keteleeria davidiana* [NC_011930], *Pset. wilsoniana* [AB601120], *L. decidua* [AB501189], *Pic. morrisonicola* [AB480556], *Cathaya argyrophylla* [AB547400], and *Pinus thunbergii* [NC_001631]) were submitted to Mulan (http://mulan.dcode.org/) to conduct a genome-wide local multiple alignment with the cpDNA of *Cedrus* used as the outgroup. Outputs of the alignment profiles were produced using the categories of "summary conservation" and "standard stacked-pairwise" for "visualization type."

Table 1
Summary of cpDNA Forms Found in 53 Pinaceae Species

Subfamily	Species	CpDNA Form	Accession	References
Abietoideae	*Abies alba*	B	—	
	A. concolor	A	—	
	A. firma	A, B	—	Tsumura et al. (2000)
	A. homolepis	A, B	—	Tsumura et al. (2000)
	A. koreana	A	—	
	A. magnifica	A	—	
	A. mariesii	A, B	—	Tsumura et al. (2000)
	A. nordmanniana	A	—	
	A. religiosa	B	—	
	A. sachalinensis	A, B	—	Tsumura et al. (2000)
	A. veitchii	A, B	—	Tsumura et al. (2000)
	Cedrus deodara	A	AB480043	Lin et al. (2010)
	Tsuga canadensis	A	—	
	T. chinesis	A	—	
	T. diversifolia	A	—	Tsumura et al. (2000)
	T. heterophylla	B	—	
	T. sieboldii	A	—	Tsumura et al. (2000)
	T. mertensiana	A	—	
	Keteleeria calcarea	A	—	
	K. davidiana	A	NC_011930	Wu et al. (2009)
	K. evelyniana	A	—	
	Pseudolarix kaempferi	A	—	
Laricoideae	*Cathaya argyrophylla*	A	AB547400	Lin et al. (2010)
	Larix decidua	C	AB501189	
	L. gmelinii	C	—	
	L. griffithiana	C	—	
	L. kaempferi	C	—	
	Pseudotsuga macrocarpa[a]	A, B	—	
	Ps. menziesii	B	—	Strauss et al. (1988)
	Ps. wilsoniana[a]	A, B	AB601120	
Piceoideae	*Picea abies*	A	—	
	Pic. glauca	A	—	
	Pic. glehnii	A	—	
	Pic. morrisonicola	A	AB480556	
	Pic. obovata	A	—	
	Pic. omorika	A	—	
	Pic. orientalis	A	—	
	Pic. schrenkiana	A	—	
	Pic. sitchensis	P	NC_011152	Cronn et al. (2008)
	Pic. rubens	A	—	
Pinoideae	*Pinus contorta*	P	NC_011153	Cronn et al. (2008)
	Pin. elliottii	C	—	
	Pin. gerardiana	P	NC_011154	Cronn et al. (2008)
	Pin. koraiensis	P	NC_004677	Noh et al. (2003)
	Pin. krempfii	P	NC_011155	Cronn et al. (2008)
	Pin. lambertiana	P	NC_011156	Cronn et al. (2008)
	Pin. longaeva	P	NC_011157	Cronn et al. (2008)
	Pin. massoniana	B	—	
	Pin. monophylla	P	NC_011158	Cronn et al. (2008)
	Pin. nelsonii	P	NC_011159	Cronn et al. (2008)
	Pin. radiata	P	—	Strauss et al. (1988)
	Pin. thunbergii	P	NC_001631	Wakasugi et al. (1994)
	Pin. wallichiana	P	—	

[a] Two isomeric cpDNAs within an individual.

Results

Features and Mutation Hotspots of Pinaceae CpDNAs

The cpDNAs of *L. decidua* (AB501189), *Pic. morrisonicola* (AB480556), and *Ps. wilsoniana* (AB601120) are circular and 122,474, 124,168, and 122,513 bp long, respectively (supplementary fig. 1, Supplementary Material online). Their IRs represent only 0.36% (436 bp), 0.35% (440 bp), and 0.28% (345 bp) of their respective cpDNA lengths and contain only 2 genes, *ψpsbA* (i.e., *3′psbA*) and *trnI-CAU*. Four previously elucidated cpDNAs of Pinaceae, one representative of each genus (*Pinus*: *Pin. thunbergii* [Wakasugi et al. 1994], *Keteleeria*: *K. davidiana* [Wu et al. 2009], *Cathaya*: *Ca. argyrophylla*, and *Cedrus*: *Ce. deodara* [Lin et al. 2010]), were included for cpDNA-wide multiple DNA alignments and comparisons. The cpDNA of *Cedrus* was used as the reference because the genus was resolved as the basal-most clade of the subfamily Abietoideae and dated to be the oldest genus in the Pinaceae (ca. 210 Ma; Lin et al. 2010). We detected four mutation hotspots with low sequence conservation (supplementary fig. 2A, Supplementary Material online). The hotspots include the intergenic respective spacer between *trnE-UUC* and *trnT-GGU* and the three regions containing the pseudogenes *ψndhH-E* cluster, *ψndhD*, and *ψycf2*. The *trnE-UUC–trnT-GGU* spacer is also a hotspot for rearrangements and is highly variable in length because of the presence/absence of repeats (see the section of Three Types of Pinaceae-Specific Repeats Are Hotspots for CpDNA Rearrangements) (supplementary fig. 2B, Supplementary Material online). Intriguingly, the above three pseudogene-containing sequences show various degrees of degradation. Loss of all *ndh* genes and one of the two copies of *ycf2* genes occurred in the common ancestor of Pinaceae cpDNAs (Wu et al. 2007), implying that these pseudogenes have been retained for at least 225 Ma (Miller 1999).

Four Distinct CpDNA Forms in Pinaceae

Figure 1 presents dot-plot comparisons between the cpDNA organizations of *Cedrus* and one representative from each of the six other Pinaceae genera. The cpDNAs of *Cedrus*, *Keteleeria*, *Picea*, and *Cathaya* have collinear organizations, except that a ~12-kb fragment (from *trnV-GAC* to *ycf2*) is missing in the cpDNA of *Cathaya* (Lin et al. 2010). However, the cpDNA organizations of *Pseudotsuga*, *Larix*, and *Pinus* differ from those of the above four in the region between *trnR-UCU* and *ψtrnG-GCC*, which comprises two fragments flanked by *trnR-UCU* and *trnE-UUC* (~20 kb; hereafter designated as F1) and by *ψrps4* and *ψtrnG-GCC* (~21 kb; hereafter designated as F2), respectively (fig. 1).

From the relative locations and orientations of F1 and F2 to those of the reduced IR_A and IR_B, we recognized four distinct cpDNA forms—P, A, B, and C—in Pinaceae. The letter "P" denotes the *Pinus* form, which was previously characterized in the species *Pin. radiata* (Strauss et al. 1988), and the A and B forms were defined by Tsumura et al. (2000), although Strauss et al. (1988) had previously described the cpDNA organization of *Ps. menziesii* (B form) without giving it a specific name. The P form is represented by the cpDNA of *Pinus*, which has the organization +F1 and –F2 ("+" denotes the forward strand and "–" the reverse strand). The A form, exemplified by the cpDNAs of *Cedrus*, *Keteleeria*, *Picea*, and *Cathaya*, has the +F1 and +F2 organization. In contrast, the B form, exemplified by the cpDNA of *Pseudotsuga*, has the organization of –F2 and –F1. The C form with the combined organization of +F2 and –F1 is represented by the cpDNA of *Larix*.

Constrained Diversity of the Pinaceae CpDNA Forms

There are eight possible ways to arrange the F1 and F2 fragments relative to one another and to the adjacent syntenic regions in the genome (fig. 2 A–G and P). Because only four forms have been reported (fig. 2 A–C and P), we wondered whether four other forms (D–G) exist or coexist in the Pinaceae cpDNAs. To answer this question, we first examined the structural organizations of the 22 elucidated cpDNAs of Pinaceae (table 1) and then conducted PCR assays to examine cpDNA organizations of 31 additional Pinaceae species, including the three whose complete cpDNAs are first reported in this study. We designed a combination of five specific primers to diagnose the orientations of F1 and F2 (fig. 2). Figure 2*B* shows that a specific form is verified when two primer pairs simultaneously produce expected PCR products. In addition, we sequenced some PCR products (e.g., *Abies concolor* [AB608752], *Ps. wilsoniana* [AB608750], and *Tsuga canadensis* [AB608751]) to confirm whether the amplified products were indeed the correct targets. For each Pinaceae species, we used eight different primer pairs independently to verify the presence of isoforms in an individual.

Table 1 summarizes the distribution of the cpDNA forms in 53 Pinaceae species, which represents all the ten Pinaceous genera and four subfamilies. The results indicated the followings: 1) we detected only four cpDNA forms (A–C and P); 2) the subfamily Abietoideae contains more than half of the Pinaceae genera (six of ten genera), but its cpDNA forms show the lowest diversity; 3) the monotypic subfamily Pinoideae has the most diversified cpDNA forms; 4) *Ps. macrocarpa* and *Ps. wilsoniana* have isomeric cpDNAs within individuals; finally, 5) with the exception of *Pseudotsuga*, the cpDNA of an individual likely has only one form, and different individuals or populations of the same species might have isoforms (e.g., *Abies firma*). Therefore, the absence of the four hypothetically potential forms (D–G) suggests constrained diversity of the Pinaceae cpDNA forms.

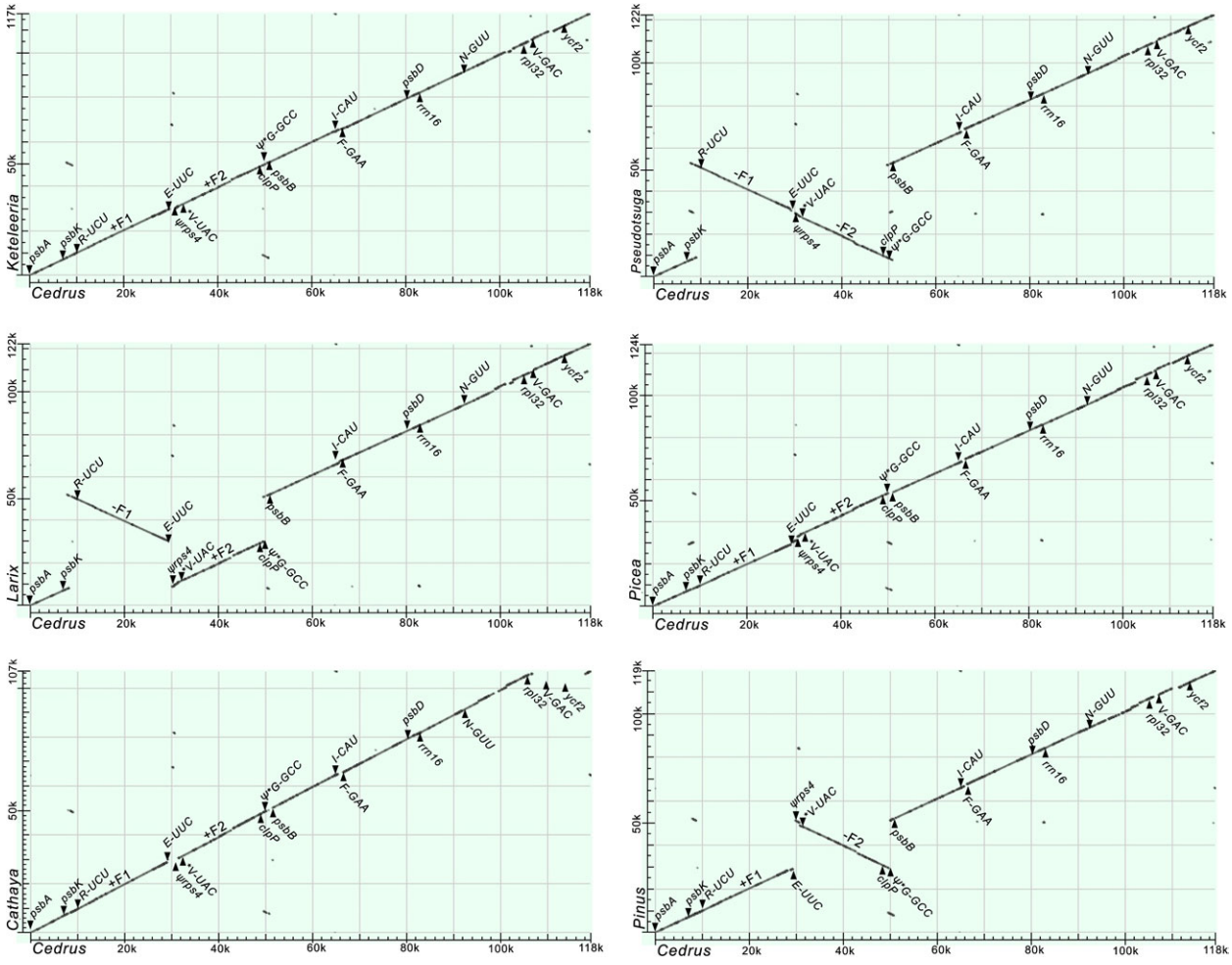


Fig. 1.—CpDNA dot-plot analyses of Pinaceae genera. The *Cedrus* cpDNA is used as the reference. A positive slope denotes that the compared two sequences (horizontal and vertical axes) are matched in the same orientations, whereas a negative one indicates that the two sequences can be aligned but with opposite orientations. Labeled genes are based on their positions in the *Cedrus* cpDNA.

The A Form Is Probably the Most Primitive in Pinaceae

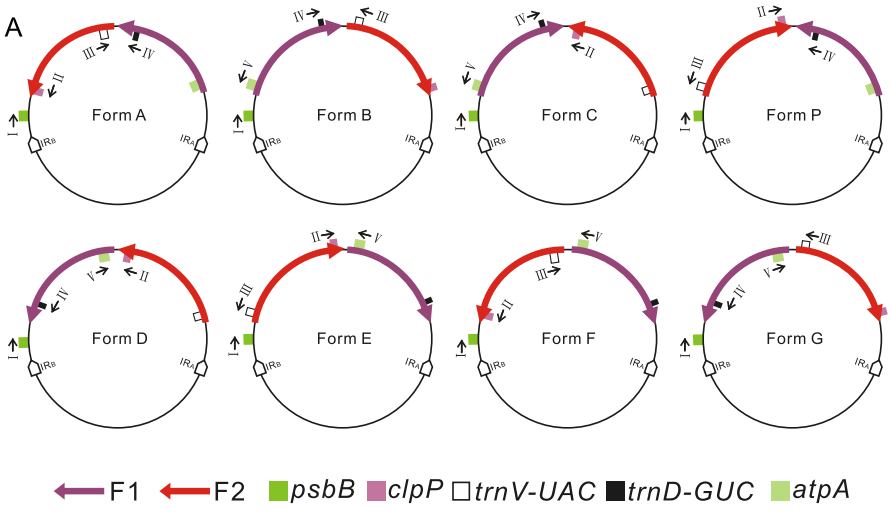
To unravel which of the four detected Pinaceae cpDNA forms represents the most ancient organization, the cpDNA of *Cycas taitungensis* was used as the outgroup. The only available cpDNA of a non-Pinaceae conifer, *C. japonica* of Cupressaceae (Hirao et al. 2008), is too highly arranged to relate with the Pinaceae cpDNAs. The A form and *Cycas* cpDNA show the same relative locations and orientations for the F1 and F2 fragments (supplementary fig. 3, Supplementary Material online), which suggests that the A form is probably more ancient than the other three.

Three Types of Pinaceae-Specific Repeats Are Hotspots for CpDNA Rearrangements

We compared the boundaries of F1 and F2 among the seven elucidated Pinaceae cpDNAs and identified three types of Pinaceae-specific repeats (designated as types 1, 2, and 3 repeats). The three types are each alignable among variants from sampled species (supplementary file 1, Supplementary Material online). Notably, although these repeats are generally found in the Pinaceae cpDNAs, their sizes are variable, with dynamic conversions between their own direct and inverted copies depending on the cpDNA forms (supplementary table 1, Supplementary Material online).

Figure 3 depicts the relative locations of the three types of Pinaceae-specific repeats. Note that all the sampled cpDNAs have two copies of type 1 repeats. The two copies are inverted relative to each other and flank the region encompassing the F1 and F2 fragments. Type 1 repeats vary from 474 to 1,335 bp and usually contain a *trnS-GCU* gene, except for those of *Picea*. Type 2 repeats were found in the cpDNAs of *Keteleeria* (three copies), *Larix* (two copies), and *Pinus* (two copies). All type 2 repeats contain the pseudogene ψ*trnG-GCC* (5′*trnG-GCC*), and each of the sampled taxa has one copy of the repeats at the junction of the F1 and F2 fragments. The cpDNAs of *Picea* and *Pseudotsuga* contain two and three copies of type

A



B

Primer pairs	CpDNA form							
	A	B	C	P	D	E	F	G
psbB-R (I) + clpP-R (II)	+	−	−	−	−	−	+	−
trnV-F (III) + trnD-R (IV)	+	+	−	−	−	−	−	−
psbB-R (I) + trnV-F (III)	−	−	−	+	−	+	−	−
clpP-R (II) + trnD-R (IV)	−	−	+	+	−	−	−	−
psbB-R (I) + atpA-F (V)	−	+	+	−	−	−	−	−
clpP-R (II) + atpA-F (V)	−	−	−	−	+	+	−	−
trnV-F (III) + atpA-F (V)	−	−	−	−	−	−	+	+
psbB-R (I) + trnD-R (IV)	−	−	−	−	+	−	−	+

Fig. 2.—Experimental verification of cpDNA forms in Pinaceae. (*A*) Sketches of eight different cpDNA forms. The reduced IR_A and IR_B are denoted by blank arrow boxes. The purple and red curved lines represent the F1 (from *trnR-UCU* to *trnE-UUC*) and F2 fragments (from *ψrps4* to *ψtrnG-GCC*), respectively. The arrows denote their relative orientations. Primers (small open arrows) designed to verify different forms are labeled. Combinations of primer pairs are shown in the table below. (*B*) The primers for PCR verification. Two primer pairs were used to determine a specific cpDNA form. "+" indicates that the primers can produce expected amplicons, whereas "−" indicates that the primers cannot produce expected ones. For example, if the two primer pairs, psbB-R (I) + clpP-R (II) and trnV-F (III) + trnD-R (IV), can simultaneously produce positive fragments, the examined cpDNA may be an "A" form.

3 repeats, respectively. Type 3 repeats contain the genes *psbI* and *trnS-GCU*, and each taxon has a single copy of the repeat between F1 and F2. In terms of size, type 3 repeats (mean size: 608 ± 73 bp) are longer than type 2 repeats (mean size: 151 ± 30 bp). Because all the 3 repeat types commonly reside near the F1 and F2 fragments, they might be potential hotspots for rearrangements.

The Highly Reduced IRs Have Lost HR Ability

We designed a simple assay using PCR methods to easily determine whether the reduced IR_A and IR_B (size <1 kb) of Pinaceae cpDNAs can still conduct IR-mediated HR (flip-flop recombination) similar to those of other plants (Palmer 1983; Cattolico et al. 2008). The two isomeric cpDNAs should have opposite orientations of single-copy regions to each other because of the IR-mediated HR. To ensure whether an IR-mediated isomeric cpDNA exists in a Pinaceae cpDNA species, we designed two primer pairs (viz. trnK-R + ycf2-F for IR_A, and rpl2-R + trnF-F for IR_B) to amplify two specific fragments across the reduced IRs (supplementary fig. 4, Supplementary Material online). If the reversed primer pairs of the above two pairs (viz. rpl2-R + ycf2-F and trnK-R + trnF-F) yield expected PCR products for a particular genomic DNA species, an IR-mediated isomeric cpDNA should exist inside the chloroplasts of the sampled species. However, our PCR assays showed that the above-mentioned reversed primer pairs did not yield any products from the 31 sampled species (data not shown). The common absence of an IR-mediated isomeric cpDNA in IR-reduced cpDNAs of Pinaceae strongly suggests that the reduced IRs of Pinaceae have lost the ability to mediate HR.

Discussion

Evolutionary Significance of the Diversified Pinaceae CpDNAs

We conclude that Pinaceae cpDNAs have four distinct forms. Previously, B (*Ps. menziesii*) and P (*Pin. radiata*) forms were mapped by Strauss et al. (1988). Later, Tsumura et al.

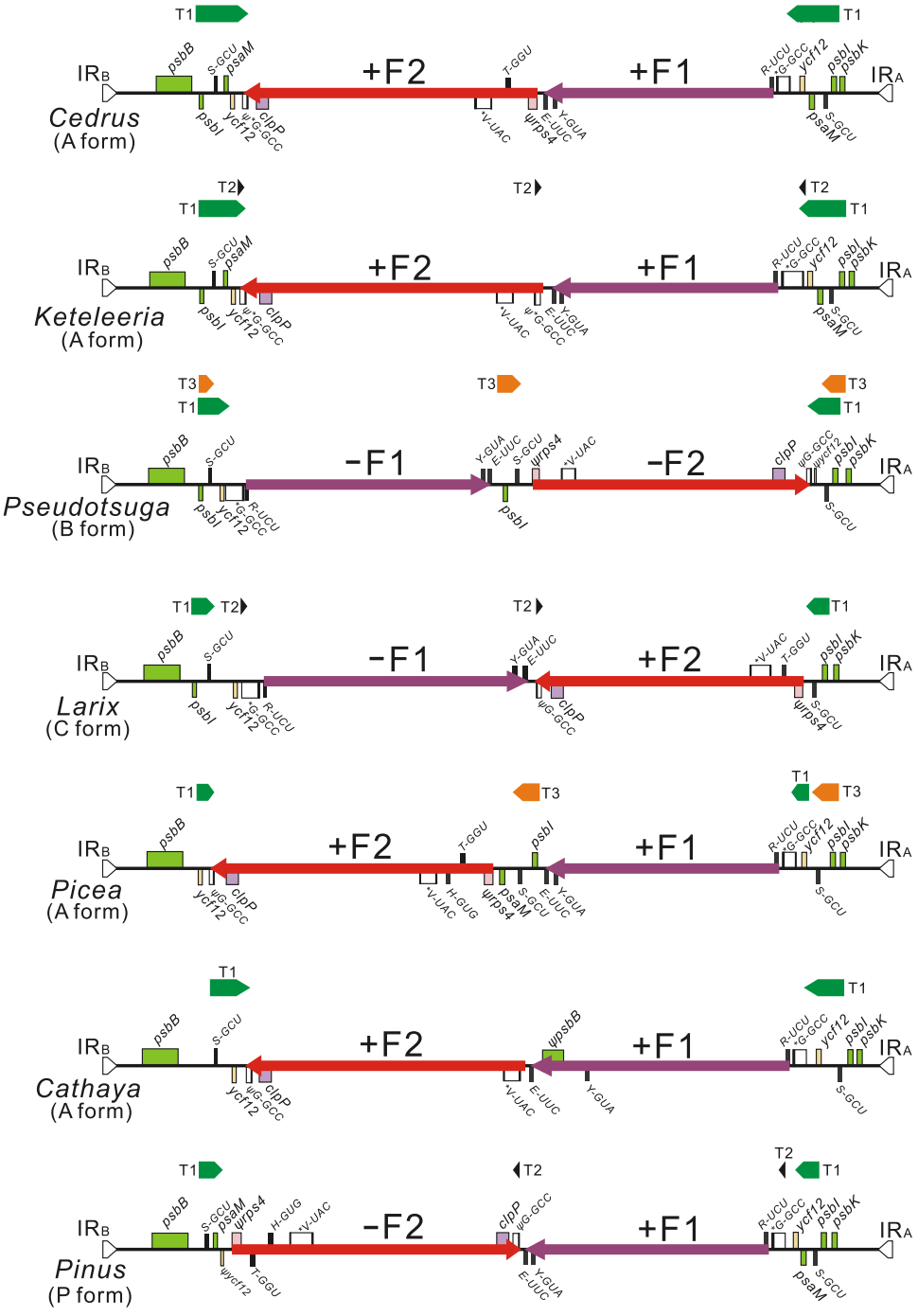


Fig. 3.—Conserved repeats located near the boundaries of the rearranged fragments, F1 (purple arrows) and F2 (red arrows). Genera are arranged on the basis of the phylogenetic frame of Lin et al. (2010). The relative locations and orientations of F1 and F2 for each genus were labeled with "+" or "−" based on the dot-plot analyses in the fig. 1. The type 1, 2, and 3 repeats are abbreviated as T1, T2, and T3 and labeled above the cpDNA of each sampled genus. The sizes of repeats are proportional to each other.

(2000) recognized A and B forms in different populations of five *Abies* and two *Tsuga* species in Japan. In the present study, we discovered the fourth form, C, represented by the cpDNA of *Larix*. Our comparisons further indicate that in the Pinaceae genera, cpDNA organizations are not correlated with phylogenetic relationships of genera. For instance, the cpDNA organization of *Picea* (A form) is more similar to that of *Cedrus* (A form) than to that of *Pinus* (P form), although *Picea* is phylogenetically closer to *Pinus* than to *Cedrus* (Wang et al. 2000; Gernandt et al. 2008; Lin et al.

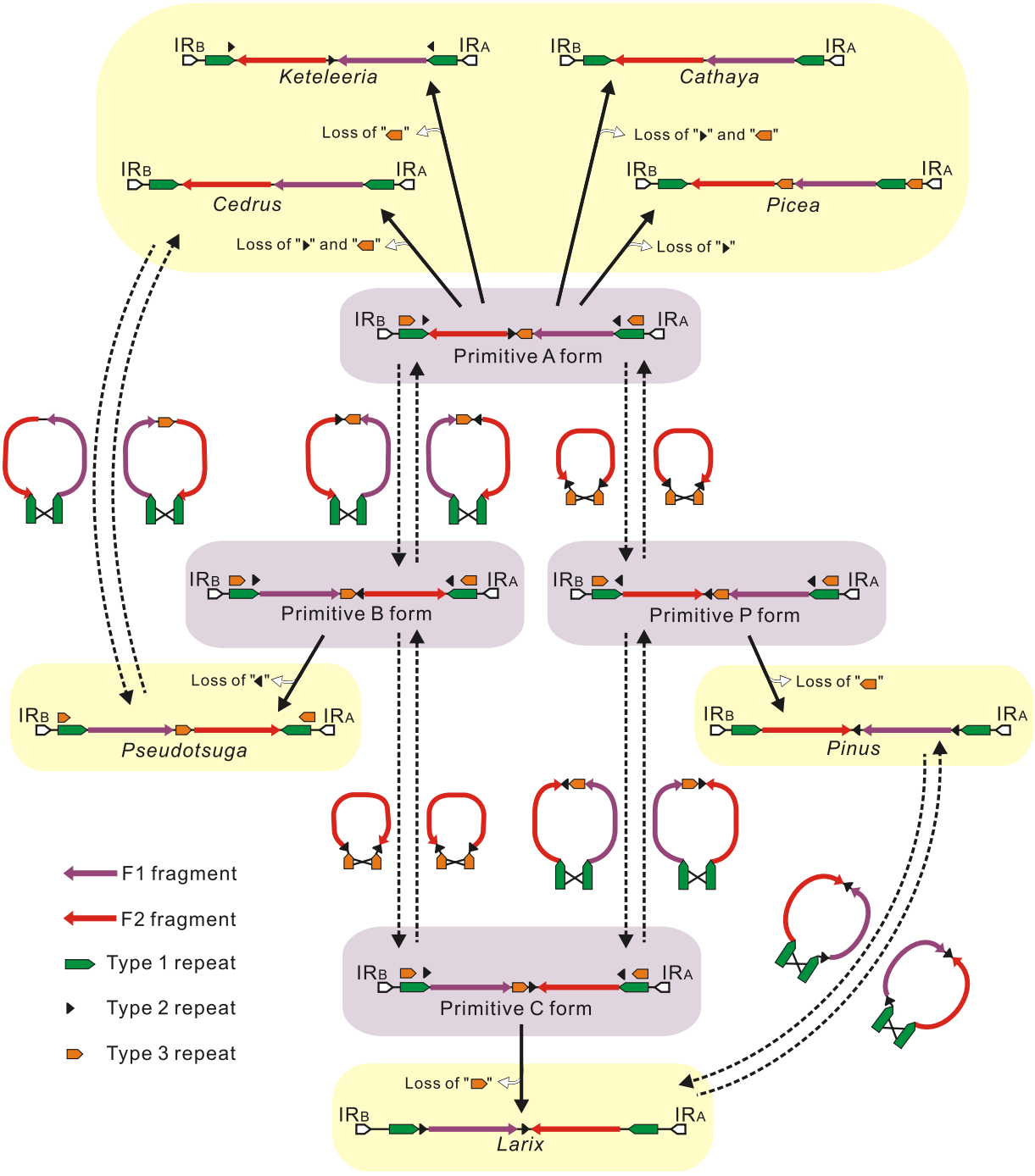


Fig. 4.—A hypothetical scenario for the formation of four distinct cpDNA forms. The extant and primitive states are highlighted with a yellow and a purple background, respectively. Conversion between two distinct forms can be achieved via hypothesized paths of HR (discontinued arrows). Specific repeats for each HR are shown along the discontinued arrows. Black arrows indicate the evolutionary direction from the primitive to extant states. White arrows along the black ones denote lost events. All arrows are not scaled.

2010). One must also consider sampling effects in a comparative study of Pinaceae cpDNAs because different intraspecific populations might possess distinct cpDNA forms (Tsumura et al. 2000). Thus, the cpDNA forms of the 53 Pinaceae species we examined might not be sufficient for depicting the full distribution spectrum because of lack of data from different populations of each sampled species.

We found that eight of the ten genera across three of the four Pinaceae subfamilies have the A form cpDNAs, which suggests that the A form is dominant. The A form might

represent a symplesiomorphic character or confer some selective advantages over the three other forms. Because an unbiased distribution of both A and B forms was previously reported in *Abies* and *Tsuga* (Tsumura et al. 2000), evolution might not have favored a specific form; whereas the symplesiomorphy of A form is consistent with the structural comparison of cpDNAs between Pinaceae and *Cycas*, showing that the A form appears to be ancestral.

A Hypothetical Scenario for the Evolution of Four Pinaceae CpDNA Forms

Although P and B and A forms were characterized by Strauss et al. (1988) and Tsumura et al. (2000), respectively, their evolutionary relatedness remained unknown. By adding the newly recognized C form, we were able to propose an evolutionary scenario for the four Pinaceae cpDNA forms based on four clues: 1) the A form is the most primitive, 2) rearrangements are constrained to the F1 and F2 fragments, 3) conserved short repeats reside at the boundaries of rearrangements, and 4) only A–C and P forms, not the putative D–G forms, exist in Pinaceae. Previously, the type 1, 2, and 3 repeats were reported to be hotspots for rearrangements in the cpDNAs of *Pseudotsuga* (Hipkins et al. 1994), *Pinus* (Wakasugi et al. 1994), and *Abies* (Tsumura et al. 2000). Indeed, short repeats have been considered to play a key role in cpDNA rearrangements of many angiosperms (e.g., *Aegilops* [Ogihara et al. 1988], *Trifolium* [Milligan et al. 1989], *Pelargonium* [Chumley et al. 2006], and *Trachelium* [Haberle et al. 2008]) because repetitive sequences can promote HR (Crouse et al. 1986; Ogihara et al. 1988; Kawata et al. 1997).

In cpDNAs, both inter- and intramolecular HR have been reported (Day and Madesis 2007). However, the Pinaceae cpDNA forms do not appear to undergo intermolecular HR because we did not find any conjugation of two unrelated segments similar to what was found in a chimeric pseudogene of rice cpDNA (Hiratsuka et al. 1989). In addition, recombination of two cpDNAs would generate a dimeric cpDNA (Kolodner and Tewari 1979), which requires a large deletion to recover a monomer, as was proposed by Hiratsuka et al. (1989) in a model for the rice cpDNA. Considering intermolecular HR, the Pinaceae cpDNAs would have to experience at least three independent large deletions to yield three distinct forms from the most primitive one. This scenario appears to violate the parsimony principle.

Figure 4 illustrates a hypothetical scenario for the formation of A–C and P forms based on intramolecular HR. Each of the three types of repeats is present in at least two Pinaceae genera, which led us to assume that these three Pinaceae-specific repeats are symplesiomorphic characters in the Pinaceae cpDNAs. All of possible HR between direct repeats are not taken into consideration because such HR will result in deletion of a large fragment, as well as loss of many essential genes. The ancestral form, A, can convert to the B or P form via HR mediated by type 1 or type 3 repeats, respectively, whereas the C form can be derived from the B or P form by type 3 repeat- or type 1 repeat-mediated HR. Theoretically, all of above HR are reversible. Intriguingly, conversion of A and B forms to the putative F and G forms is likely achieved by type 2 repeat-mediated HR. However, we did not detect any of the latter forms in our experiments, which suggests that type 2 repeats might lack HR-mediated ability. Because a length of ~200 bp is considered necessary for efficient HR (Day and Madesis 2007) and the average length of type 2 repeats is only 151 ± 30 bp, such a short length may disable type 2 repeats to mediate HR.

Evolutionary Trends of Pinaceae CpDNAs

Although the three types of Pinaceae-specific repeats offer reliable clues to unravel the formation mechanism for the four Pinaceae cpDNA forms, they have different selective pressures. All Pinaceae genera have retained the type 1 repeats, but only three and two sampled genera had the type 2 and type 3 repeats, respectively. In addition, the type 2 repeats might not be able to activate HR as mentioned above. Loss of the type 3 repeats interrupts the conversion between A and P forms and between B and C forms (fig. 4). Of note, only *Picea* and *Pinus* (two more recently diverged genera of the family; see Lin et al. 2010) have both A and P forms and both B and C forms, respectively. Therefore, the Pinaceae cpDNAs have evolved toward the conversion between A and B forms and between C and P forms. In contrast, the conversion between A and P forms and between B and C forms might occur rarely.

The Effect of IR Reduction in Pinaceae

To date, the function of IRs remains uncertain. Previously, Palmer and Thompson (1981) suggested that IRs may best be viewed as an evolutionary relic. Later, from information on the increased diversity of cpDNA organizations in IR-lacking legumes, Palmer and Thompson (1982) suggested that IRs might stabilize cpDNA organizations. Strauss et al. (1988) supported the view of Palmer and Thompson (1982) and proposed that in the Pinaceae cpDNAs, rearrangements have taken place after IR reduction. However, these authors did not show how IR reduction led to rearranged cpDNA organizations in Pinaceae. IRs have been reported to mediate intramolecular HR that yields equal populations of two isomeric cpDNAs with opposite orientations in the single-copy regions (Bohnert and Loffelhardt 1982; Palmer 1983; Stein et al. 1986; Cattolico et al. 2008).

Our PCR assays suggest that in Pinaceae, the reduced IRs might have lost the ability to activate HR. Because their mean size is 362 ± 101 bp (Lin et al. 2010), about 0.4 and 0.6 times the size of the type 1 and 3 repeats,

respectively, they might be less competitive for HR than the type 1 and type 3 repeats. Combining the data of Tsumura et al. (2000) and ours, we propose that in Pinaceae, IR reduction may lead to uniformity of the cpDNA organizations within an individual or a population. Genetic uniformity is widely accepted to be an evolutionary disadvantage that reduces the ability to protect against environmental stress. Although almost all of our surveyed Pinaceae species did not show active HR, we did find two isomeric cpDNAs that likely resulted from the type 1 repeat-mediated HR, and they coexist within individuals of both *Ps. macrocarpa* and *Ps. wilsoniana*. Tsumura et al. (2000) reported that isomeric cpDNAs are copresent in individuals of both *Abies* and *Tsuga*, although their PCR signals were weak. Therefore, we conclude that evolution has endowed Pinaceae with short repeats, specifically the type 1 repeats, which can complement the reduced IRs and also increase the diversity of cpDNA forms.

Conclusions

Our analyses revealed that in the Pinaceae cpDNAs, rearrangements of two specific fragments generate four distinct cpDNA forms. These forms are common in the ten Pinaceae genera. We discovered that three major types of Pinaceae-specific repeats (types 1–3), situated near the boundaries of the rearranged fragments, are syntenic in the cpDNAs of Pinaceae. Two of these repeats (type 1 [949 ± 343 bp] and type 3 [608 ± 73 bp]) may serve as "hotspots" for rearrangements via HR. The type 2 repeats are likely inactive for HR because they are relatively short (151 ± 30 bp). We present a hypothetical model for the structural evolution of Pinaceae cpDNAs. Our PCR analyses suggest that the highly reduced IRs might have lost the ability to induce HR and been replaced by the type 1 and 3 repeats when the common ancestor of the Pinaceae evolved approximately 225 Ma (Miller 1999).

Supplementary Material

Supplementary figures 1–4, table 1, file 1 are available at *Genome Biology and Evolution online* (http://gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by research grants from the National Science Council, Taiwan (NSC972621B001003MY3) and the Biodiversity Research Center, Academia Sinica to S.M.C. We thank Yi-Ming Chen for the materials of *Cathaya*, *Cedrus*, and *Pseudolarix*, and Shu-Mei Liu, Shu-Jen Chou, and Mei-Jane Fang for the help with DNA shearing and sequencing. We are grateful to the two anonymous reviewers for their critical reading and helpful suggestions in improving the manuscript.

Literature Cited

Bohnert HJ, Loffelhardt W. 1982. Cyanelle DNA from *Cyanophora paradoxa* exists in two forms due to intramolecular recombination. FEBS Lett. 150:403–406.

Cai Z, et al. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol. 67: 696–704.

Cattolico RA, et al. 2008. Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. BMC Genomics. 9:211.

Chumley TW, et al. 2006. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 23:2175–2190.

Cronn R, et al. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res. 36:e122.

Crouse EJ, et al. 1986. Divergence of chloroplast gene organization in three legumes. Plant Mol Biol. 7:143–150.

Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: Bock R, editor. Cell and molecular biology of plastids. Topics in current genetics. Vol. 19. Heidelberg (Germany): Springer. p. 65–119.

Gernandt DS, et al. 2008. Use of simultaneous analyses to guide fossilbased calibrations of Pinaceae phylogeny. Int J Plant Sci. 169:1086–1099.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 28:583–600.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol. 66:350–361.

Hipkins VD, Krutovskii KV, Strauss SH. 1994. Organelle genomes in conifers: structure, evolution, and diversity. For Genet. 1:179–189.

Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8:70.

Hiratsuka J, et al. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet. 217:185–194.

Kawata M, Harada T, Shimamoto Y, Oono K, Takaiwa F. 1997. Short inverted repeats function as hotspots of intermolecular recombination giving rise to oligomers of deleted plastid DNAs (ptDNAs). Curr Genet. 31:179–184.

Kim KJ, Choi KS, Jansen RK. 2005. Two chloroplast DNA inversions originated simultaneously during early evolution in the sunflower family. Mol Biol Evol. 22:1783–1792.

Kim KJ, Lee HL. 2005. Widespread occurrence of small inversions in the chloroplast genomes of land plants. Mol Cells. 19:104–113.

Kolodner R, Tewari KK. 1979. Inverted repeats in chloroplast DNA from higher plants. Proc Natl Acad Sci U S A. 76:41–45.

Lee HL, Jansen RK, Chumley TW, Kim KJ. 2007. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. Mol Biol Evol. 24: 1161–1180.

Lin CP, Huang JP, Wu CS, Hsu CY, Chaw SM. 2010. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. Genome Biol Evol. 2:504–517.

McCoy SR, Kuehl JV, Boore JL, Raubeson LA. 2008. The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. BMC Evol Biol. 8:130.

Miller CN. 1999. Implications of fossil conifers for the phylogenetic relationships of living families. Bot Rev. 65:239–277.

Milligan BG, Hampton JN, Palmer JD. 1989. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol Biol Evol. 6:355–368.

Noh EW, et al. 2003. Complete nucleotide sequence of Pinus koraiensis. Direct Submission to GenBank. Accession No. NC_00467

Ogihara Y, Terachi T, Sasakuma T. 1988. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. Proc Nat Acad Sci U S A. 85:8573–8577.

Palmer JD, Thompson WF. 1981. Rearrangements in the chloroplast genomes of mung bean and pea. Proc Natl Acad Sci U S A. 78:5533–5537.

Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29:537–550.

Palmer JD. 1983. Chloroplast DNA exists in 2 orientations. Nature 301:92–93.

Palmer JD. 1991. Plastid chromosomes: structure and evolution. In: Bogorad L, editor. Molecular biology of plastids. San Diego (CA): Academic Press. p. 5–53.

Pombert JF, Otis C, Lemieux C, Turmel M. 2005. The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. Mol Biol Evol. 22:1903–1918.

Pombert JF, Lemieux C, Turmel M. 2006. The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. BMC Biol. 4:3.

Raubeson LA, Jansen RK. 1992. A rare chloroplast DNA structural mutation is shared by all conifers. Biochem Syst Ecol. 20:17–24.

Saski C, et al. 2005. Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. Plant Mol Biol. 59:309–322.

Shimada H, Sugiura M. 1989. Pseudogenes and short repeated sequences in the rice chloroplast genome. Curr Genet. 164:293–301.

Stein DB, Palmer JD, Thompson WF. 1986. Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus *Osmunda*. Curr Genet. 10:835–841.

Stewart CN Jr, Via LE. 1993. A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. BioTechniques 14:748–751.

Strauss SH, Palmer JD, Howe GT, Doerksen AH. 1988. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc Natl Acad Sci U S A. 85:3898–3902.

Tomioka N, Sugiura M. 1983. The complete nucleotide sequence of a 16s ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. Mol Gen Genet. 191:46–50.

Tsai CH, Strauss SH. 1989. Dispersed repetitive sequences in the chloroplast genome of Douglas-fir. Curr Genet. 16:211–218.

Tsudzuki J, et al. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ, trnK, psbA, trnI* and *trnH* and the absence of *rps16*. Mol Gen Genet. 232:206–214.

Tsumura Y, Suyama Y, Yoshimura K. 2000. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. Mol Biol Evol. 17:1302–1312.

Turmel M, Otis C, Lemieux C. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. Proc Natl Acad Sci U S A. 96:10248–10253.

Wakasugi T, et al. 1994. Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc Natl Acad Sci U S A. 91:9794–9798.

Wang XQ, Tank DC, Sang T. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. Mol Biol Evol. 17:773–781.

Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. Mol Phylogenet Evol. 52:115–124.

Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. Mol Biol Evol. 24:1366–1379.

Associate editor: Yves Van De Peer